Accounting for extrinsic variability in the estimation of stochastic rate constants

Koeppl, Heinz ; Zechner, Christoph ; Ganguly, Arnab ; Pelet, Serge ; Peter, Matthias

Abstract: Single-cell recordings of transcriptional and post-transcriptional processes reveal the inherent stochasticity of cellular events. However, to a large extent, the observed variability in isogenic cell populations is due to extrinsic factors, such as difference in expression capacity, cell volume and cell cycle stage—to name a few. Thus, such experimental data represents a convolution of effects from stochastic kinetics and extrinsic noise sources. Recent parameter inference schemes for single-cell data just account for variability because of molecular noise. Here, we present a Bayesian inference scheme that deconvolutes the two sources of variability and enables us to obtain optimal estimates of stochastic rate constants of low copy-number events and extract statistical information about cell-to-cell variability. In contrast to previous attempts, we model extrinsic noise by a variability in the abundance of mass-conserved species, rather than a variability in kinetic parameters. We apply the scheme to a simple model of the osmostress-induced transcriptional activation in budding yeast.

DOI: https://doi.org/10.1002/rnc.2804

INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust. Nonlinear Control 0000; **00**:1–19
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rnc

Accounting for Extrinsic Variability in the Estimation of Stochastic Rate Constants

Heinz Koeppl[1]*, Christoph Zechner[1], Arnab Ganguly[1], Serge Pelet[2] and Matthias Peter[2]

[1]*ETH Zürich, D-ITET, BISON Group, Automatic Control Lab, Physikstr. 3, 8092 Zurich, Switzerland*
[2]*ETH Zürich, D-BIOL, Institute for Biochemistry, Schafmattstr. 18, 8093 Zurich, Switzerland*

SUMMARY

Single-cell recordings of transcriptional and post-transcriptional processes reveal the inherent stochasticity of cellular events. However, to a large extent the observed variability in isogenic cell populations is due to extrinsic factors, such as difference in expression capacity, cell volume and cell cycle stage - to name a few. Thus, such experimental data represents a convolution of effects from stochastic kinetics and extrinsic noise sources. Recent parameter inference schemes for single-cell data just account for variability due to molecular noise. Here we present a Bayesian inference scheme which de-convolutes the two sources of variability and enables us to obtain optimal estimates of stochastic rate constants of low copy-number events and extract statistical information about cell-to-cell variability. In contrast to previous attempts, we model extrinsic noise by a variability in the abundance of mass-conserved species, rather than a variability in kinetic parameters. We apply the scheme to a simple model of the osmo-stress induced transcriptional activation in budding yeast. Copyright © 0000 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: Cell-to-Cell Variability, Stochastic Chemical Kinetics, Mass-Conservation, Bayesian Estimation, MAPK Hog1 Signaling Pathway

1. INTRODUCTION

Our current understanding of dynamic processes inside the single cell is mostly derived from ensemble measurements. That is, most biochemical methods such as Western blotting rely on the analysis of large samples of fractionated cells [1]. Even though it is acknowledged that a population of isogenic cells exhibits considerable phenotypic heterogeneity, the working hypothesis is that the ensemble mean dynamics is a good representative of the single-cell dynamics. Such measured heterogeneity can arise due to non-synchronized cell-cycle stage [1], intrinsic molecular noise [2, 3] and effects thereof (e.g. difference in expression capacity), variations in local growth conditions and so forth. Single-cell technologies, like fluorescent activated cell sorting (FACS) and fluorescent-based microscopy and spectroscopy reveal that heterogeneity often does not show uni-modal characteristics [4, 5] – questioning the relevance of mean dynamics. That is to say, the population consists of several dominant subpopulations, each having its incremental variation around it. Take for instance a bimodal cytometric expression profile, as encountered in the case study discussed subsequently [6]. The two subpopulations correspond to cells that *have* or *have not* activated the stress-response at a certain stress level. Under the afore working hypothesis one erroneously concludes that cells show a graded response, i.e. they activated their stress-response to

H. Koeppl and C. Zechner have equally contributed.
*Correspondence to: E-mail: koepplh@ethz.ch

an intermediate level. In particular, for dynamic computational modeling that aims at capturing the underlying molecular mechanism within the single cell, such conclusions can be very misleading. Moreover, timing aspects that are crucial to develop a mechanistic understanding are generally lost in ensemble averages.

Single-cell techniques give access to the cell-to-cell variability and open up the possibility to determine its causes. In order to do so, one needs multi-variate data, for instance multiple channels from fluorescent microscopy or cytometry. With an ingenuous experiment [3], two qualitatively different contributions to this variability were singled out. One being the *intrinsic noise*, is due to the stochastic nature of chemical kinetics [7]. It gives variability in the same way as every sample path drawn from a Markov chain looks different. The remaining contribution – coined *extrinsic noise* – characterizes what is normally referred to as heterogeneity in the cell population, such as variations in total protein concentration, in cell volume and in the spatial position of regulatory molecules. A complementary analysis is discussed in [8], where contributions are separated with respect to signaling and expression components. Others found that the encountered expression noise is gene-specific and does not correlate well with the overall expression capacity of a cell [9]. Moreover, experimental advances, such as single-molecule techniques favor the further discrimination of different sources of variability [10]. Below the sensitivity of fluorescent-proteins-based techniques, single molecules approaches, such as mRNA detection using the MS2 bacteriophage, or fluorescence *in vivo* Hybridization (FIVH) have been developed to actually count single mRNA copies within a cell [11].

A key challenge in systems biology is to use those experimental data to infer mechanisms and finally computational models of cellular processes. A particular instance of these problems is the inference of kinetic reaction rate constants. Recently, attempts have been made to use single-cell measurements to infer stochastic rate constants and to exploit the observed variability[†]. Nested Markov-chain Monte Carlo methods have been deployed to bridge between discrete observations and to sample from posteriors in a Bayesian setting [12]. Reinker *et al*. fitted a hidden Markov model to noisy data of single-mRNA measurement [13], whereas Golightly and Wilkinson proposed to infer the parameters of a diffusion approximation to the jump process to cope with the state-space-explosion [14]. Wang *et al*. propose a stochastic gradient descent scheme for maximum-likelihood parameter estimation for discretely observed stochastic kinetic models [15]. Another attempt to keep the computational complexity manageable, is made by using approximate variational inference on the jump process [16]. Approximate Bayesian computation is applied in [17] to infer kinetic rates and perform model selection. Furthermore, probability metrics [18] as cost functions are considered in order to tune a stochastic model to single-cell observations [19, 20].

All the above approaches have in common that they consider stochastic models solely based on stochastic chemical kinetics. They cannot however, be used alone to address the observed cell-to-cell variability because it is well-known that the extrinsic component is in general the dominant source of variability [3, 8]. To computationally tackle that extrinsic heterogeneity has only been done in terms differential equation models [21]. Therein, individual cells are modeled by reaction-rate equations, the parameters of which vary from cell to cell but are drawn from a common distribution. The approach can account for the extrinsic component, however it cannot incorporate any intrinsic stochasticity due to molecular noise. In this paper we propose a complementary approach that can accommodate extrinsic as well as intrinsic variability. An important conceptual difference to previous works is the following. We argue that for mechanistic models, that involve only elementary mass-action events such as binding and modification a variability of kinetic rate constants from cell to cell is to some extent unjustified. The probability of those elementary events are governed by biophysical principles and should be invariant over an isogenic population. In turn, we focus on the variability in abundance of the involved biochemical species. Note that such a variability can give rise to varying kinetic parameters for aggregated reactions. For instance, a varying total enzyme

[†]Note that the kinetic reaction rate constant is related to the stochastic rate constant by a scaling factor Ω^{n-1}, where Ω and n are the volume and the reaction arity, respectively.

DOI: 10.1002/rnc

concentration in a Michaelis-Menten aggregation will enter the parameter determining the maximal reaction velocity $v_{\max}$.

Undoubtedly, many other sources of variability exist that we do not address in this work. Moreover, we remark that we deliberately stay away from two related problems of paramount importance. That is (a) the reconstruction of complete sample paths from a discretely observed time-series and (b) the inference of states for unobserved species [22, 12]. The solution to both problems is necessary in realistic scenarios. Such solutions have been proposed and they can be incorporated into the method proposed herein, but their technicalities would occlude the main message of this paper. In the subsequent Section we illustrate our approach by a simple example.

1.1. Motivation

Consider a conversion process, defined by the reversible reaction

$$A \underset{c_2}{\overset{c_1}{\rightleftharpoons}} B. \tag{1}$$

Looking at equation (1), we can observe that when starting with a given number of molecules for species A and B, the total number of molecules $b = A + B$ will be constant for all times t[‡]. We might now ask, how extrinsic cellular noise could affect the given process. For instance, we could assume that the rate constants c_1 and c_2 vary from cell to cell, obviously resulting in different process dynamics. But depending on a model's level of abstraction, stochastic rates are more or less subject to physical laws and thus, should be insensitive against cell-specific properties such as the shape or the volume. In this case, extrinsic noise can influence the process behavior only by means of the total number of molecules b, yielding different state spaces for each of the cells as depicted in Figure 1. It is clear that if both species A and B can be measured for each cell, b can be determined given

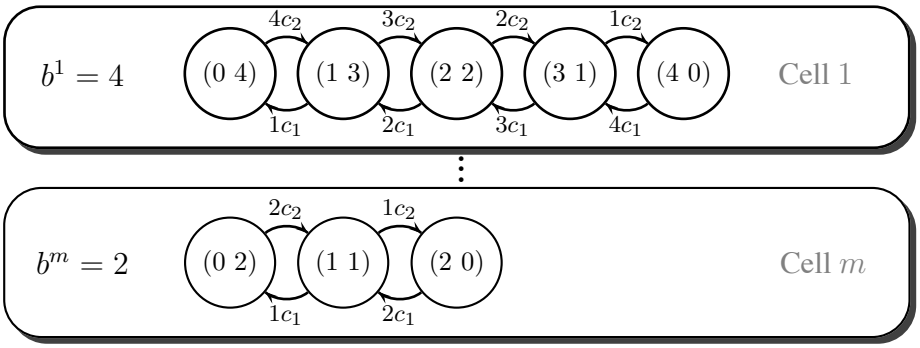


Figure 1. State diagrams for m different cells, described by the same conversion process. The graphic illustrates the heterogeneity, introduced by the variability of the conservation constant b. Each circle represents one possible state of the cell, where the two numbers denote the multiplicity of species A (left) and B (right) respectively. The arcs denote transitions between states and the labels specify the propensities given by the rates c_1 and c_2 and the corresponding multiplicities of A and B respectively. Even if the cells share the same underlying reactions and rate constants, the state space changes according to the total number of molecules $A + B = b$.

any system state. Trivially, we could then calculate statistics using each cell's conservation constant b. However, real-world measurements are always noisy, incomplete regarding the captured species as well as temporally sparse, motivating us to formulate a more general approach, which can also be used for realistic scenarios.

Assuming the simple model (1), we can first express one species by the other via b, such as for instance $B = b - A$. As we will see in Section 2, this allows us to rewrite the statistical model in

[‡]For simplicity, we do not introduce additional symbols for the species abundances but use their acronyms straight away.

DOI: 10.1002/rnc

terms of b, being the key step for quantifying the heterogeneity of multiple-cell measurements. Additionally, exploiting the presence of conservation laws eliminates the need of knowing all species involved in the model, which helps when dealing with real-world measurement data. Later in this work, we will generalize this idea for general reaction systems with arbitrary conservation laws.

The rest of the paper is structured as follows. In Section 2, we will introduce the basic concepts of the continuous time Markov chain description used in this work. Furthermore, we will account for the heterogeneity, introduced by the cell-to-cell variability. In Section 3 this will allow us to formulate an exact Bayesian inference framework for estimating the stochastic rate constants as well as inferring characteristics of the extrinsic noise. In Section 4 we will evaluate the proposed algorithms using synthetic time-course data, obtained from a simple stochastic model of stress-induced gene expression in budding yeast.

2. MODELING WITH HETEROGENEOUS CONTINUOUS TIME MARKOV CHAINS

Suppose we have a chemical reaction system with n species and v reactions. One common approach to describe the temporal evolution of chemical reaction systems is to model the system state as a *continuous-time Markov chain* (*CTMC*) $\mathbf{X}$ with state-space $\mathbb{Z}_+^n$, that is, $\mathbf{X}(t)$ is a $\mathbb{Z}_+^n$-valued random variable for all $t > 0$. If k-th reaction occurs at time t, then we update the system as $\mathbf{X}(t) = \mathbf{X}(t-) + \boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-$, where $\mathbf{X}(t-)$ denotes the state of the system just before time t and $\boldsymbol{\nu}_k^-$ and $\boldsymbol{\nu}_k^+$ represent the vector of number of molecules consumed and created in one occurrence of reaction k. Furthermore we define the stoichiometry matrix as $\mathbf{S} = \left(\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-\right)_{k\in\{1,\ldots,v\}}$. It follows that if $\mathcal{R}_k(t)$ denotes the total number of occurrences of reaction k up to time t, then

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{k=1}^{v} \mathcal{R}_k(t)(\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-).$$

Let $a_k : \mathbb{Z}_+^n \longrightarrow \mathbb{R}_+$ denote the propensity of the reaction k. The evolution of the counting process $\mathcal{R}_k$ is modeled by

$$\mathrm{P}\left(\mathcal{R}_k(t+\Delta t) - \mathcal{R}_k(t) > 0 | \mathbf{X}(t)\right) = a_k\left(\mathbf{X}(t)\right)\Delta t + o(\Delta t). \tag{2}$$

Note that the left side is just the probability of the k-th reaction firing in the small interval $[t, t+\Delta t)$. If $\mathcal{R}_k,\ k = 1,\ldots,v$ satisfy (2), then there exist independent unit Poisson processes $\xi_k,\ k = 1,\ldots,v$ (see [23, 24]) such that

$$\mathcal{R}_k(t) = \xi_k\left(\int_0^t a_k(\mathbf{X}(s))\, ds\right).$$

Hence, the state of the system is given by

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{k=1}^{v} \xi_k\left(\int_0^t a_k(\mathbf{X}(s))\, ds\right)(\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-). \tag{3}$$

As we have seen from the introductory example from Section 1.1, cell-to-cell variability introduces some kind of *heterogeneity* in the dynamics of chemical reaction systems. Thus, each cell's dynamics has to be described by an individual continuous-time Markov chain. Clearly, it would be of great interest to find a generic process formulation, which allows estimating common cell parameters as well as accounts for the extrinsic variability from cell to cell. An elegant way to do this is to write the state of the system at time t, $\mathbf{Z}(t, \mathbf{b})$, as a solution to

$$\mathbf{Z}(t, \mathbf{b}) = \mathbf{Z}(0, \mathbf{b}) + \sum_{k=1}^{v} \xi_k\left(\int_0^t \tilde{a}_k(\mathbf{Z}(s, \mathbf{b}), \mathbf{b})\, ds\right)(\boldsymbol{\nu}_k^+ - \boldsymbol{\nu}_k^-). \tag{4}$$

Prepared using rncauth.cls
DOI: 10.1002/rnc

where $\mathbf{b}$ is a parameter which varies from cell to cell. Notice that the propensities $\tilde{a}_k$ as well as the initial distribution π_0 of $\mathbf{X}$ will vary from cell to cell accordingly. In this paper we take propensities $\tilde{a}_k$ to be

$$\tilde{a}_k(\mathbf{x}, \mathbf{b}) = a_k(\mathbf{x}) 1_{\{\mathbf{Nx}=\mathbf{b}\}}, \tag{5}$$

where a_k is the propensity given by the usual mass-action kinetics and $\{\mathbf{x} \in \mathbb{Z}^n_{\geq 0} : \mathbf{Nx} = \mathbf{b}\}$ defines the state-space of the Markov chain $\mathbf{X}$. Note that relation

$$\mathbf{Nx} = \mathbf{b} \tag{6}$$

holds for arbitrary reaction systems, where we define $\mathbf{N} \in \mathbb{Z}^{l \times n}_{\geq 0}$ as a basis of the l-dimensional left null space Null $\{\mathbf{S}\}$. Then $\mathbf{b} \in \mathbb{Z}^l_{\geq 0}$ is a constant vector (i.e., *conservation constant*), specifying the position of the w-dimensional subspace $\mathbf{x}$, whereas $w = \text{Rank}\,\{\mathbf{S}\} = n - l$. In other words, $\mathbf{N}$ represents l mass-conservation laws, each of them involving a certain set of chemical species. If we furthermore restrict $\mathbf{N}$ to be the smallest positive, integer basis of Null $\{\mathbf{S}\}$, the l-th component of $\mathbf{b}$ corresponds to the total number of molecules of the l-th conservation law.

Recalling the simple example from Section 1.1, the sum of molecule numbers of A and B are conserved, which is reflected by the null space matrix $\mathbf{N} = (1, 1)$. As a consequence, the corresponding state trajectory is constrained to move along a linear subspace of dimension one.
In presence of mass-conservation laws, only w species have to be known in order to reconstruct the entire state. If we rearrange the dimensions of the system state and rewrite equation (6), we obtain

$$\mathbf{Nx} = \begin{pmatrix} \tilde{\mathbf{N}} & \bar{\mathbf{N}} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{x}} \\ \bar{\mathbf{x}} \end{pmatrix} = \mathbf{b}, \tag{7}$$

whereas $\tilde{\mathbf{N}} \in \mathbb{Z}^{l \times w}_{\geq 0}$ denotes the part of $\mathbf{N}$ corresponding to the partition $\tilde{\mathbf{x}} \in \mathbb{Z}^w_{\geq 0}$ and $\bar{\mathbf{N}} \in \mathbb{Z}^{l \times l}_{\geq 0}$ is a matrix acting on $\bar{\mathbf{x}} \in \mathbb{Z}^l_{\geq 0}$. Furthermore, we can write $\tilde{\mathbf{N}}\tilde{\mathbf{x}} + \bar{\mathbf{N}}\bar{\mathbf{x}} = \mathbf{b}$ and rearranging with respect to $\bar{\mathbf{x}}$ yields

$$\bar{\mathbf{x}} = \bar{\mathbf{N}}^{-1} \left(\mathbf{b} - \tilde{\mathbf{N}}\tilde{\mathbf{x}} \right) \equiv \mathbf{F}\left(\tilde{\mathbf{x}}, \mathbf{b}\right). \tag{8}$$

We have now found a relation, that allows us to express a set of certain species $\bar{\mathbf{x}}$ as a function of the other species $\tilde{\mathbf{x}}$ and $\mathbf{b}$. Note that $\bar{\mathbf{N}}$ must have full rank in order to be invertible, which means that in general (8) does not hold for every partitioning $\mathbf{x}$. Using the results from this section, we can now start with setting up our Bayesian inference model.

3. PARAMETER INFERENCE

3.1. Conditional Densities

Since the state-space of the Markov chain $\mathbf{X}$ is $\mathbb{Z}^n_+$, the distribution of $\mathbf{X}$ is determined by the joint distribution of the waiting time of the jumps and the reactions that happened at the jump times. The density of the latter joint distribution is required in the Bayesian analysis that we will carry out in the next section. For a random variable, the first step to find its density is to write its distribution over 'simple' sets as an integration of some appropriate function over that set. Following this approach we first calculate the joint distribution of the waiting time of the jumps and the corresponding reactions over 'cylinder' sets. These sets are precisely the ones where the above mentioned joint distribution can be computed exactly. The required density is then found out using the definition. Below, we provide rigorous derivation.

Let $\tau_1, \tau_2, \ldots$ be the jump times of the Markov chain $\mathbf{X}$, that is assuming $\tau_0 = 0$, define inductively

$$\tau_k = \inf\{s > \tau_{k-1} | \mathbf{X}(s) \neq \mathbf{X}(\tau_{k-1})\}.$$

DOI: 10.1002/rnc

Note that a jump can happen at τ_k only when a reaction occurs at that time. From (3)

$$\tau_k - \tau_{k-1} | \mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1} \ \sim Exp\left\{a(\mathbf{x}_{k-1})\right\}, \tag{9}$$

where $a(\mathbf{x}) = \sum_{j=1}^{v} a_j(\mathbf{x})$. If R_k denotes the reaction at τ_k, then

$$\mathrm{P}(R_k = j | \mathbf{X}(\tau_{k-1}) = \mathbf{x}_{k-1}) = \frac{a_j(\mathbf{x}_{k-1})}{a(\mathbf{x}_{k-1})}, \tag{10}$$

and R_k and $\tau_k - \tau_{k-1}$ are conditionally independent given $\mathbf{X}(\tau_{k-1})$. Observe that the following updating scheme holds: $\mathbf{X}(\tau_k) = \mathbf{X}(\tau_{k-1}) + \boldsymbol{\nu}^+_{R_k} - \boldsymbol{\nu}^-_{R_k}$. It is clear that the distribution of $\mathbf{X}$ will be specified by the initial distribution π_0 and by the joint distribution of $(\tau_k - \tau_{k-1}, R_k)$.

Let $d \in \mathbb{Z}_+$ be fixed. We want to find the density of $\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^d\right)$. (Observe that this means we are looking at the first d jumps of the Markov chain $\mathbf{X}$.) Note that $\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^d\right)$ is a $\mathbb{Z}_+^n \times ([0,\infty) \times \{1,\dots,v\})^d$-valued random variable.

Let $\Gamma \subset \mathbb{Z}_+^n \times ([0,\infty) \times \{1,\dots,v\})^d$ be a cylinder set of the form

$$\Gamma = \{\mathbf{x}_0\} \times \prod_{k=1}^{d} \left([\gamma_k, \eta_k] \times \{r_k\}\right) \tag{11}$$

where $0 < \gamma_k \le \eta_k < \infty$ and $\mathbf{x}_0 \in \mathbb{Z}_+^n$ and $r_k \in \{1,\dots,v\}$. Let $\mathcal{C} = \{c_1,\dots,c_v\}$ be the parameter set consisting of the stochastic rate constants. If π denotes the distribution of $\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^d\right)$ then notice that

$$\begin{aligned}
\pi(\Gamma) &\equiv \mathrm{P}\left(\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^d\right) \in \Gamma | \mathcal{C}\right) \\
&= \mathrm{P}(\tau_k - \tau_{k-1} \in [\gamma_k, \eta_k], 0 < \gamma_k < \eta_k < \infty, \ \mathbf{X}(0) = \mathbf{x}_0, \\
&\qquad R_k = r_k \ \ k = 1,2,\dots,d | \mathcal{C}) \\
&= \pi_0(\mathbf{x}_0) \prod_{j=1}^{d} \int_{\gamma_j}^{\eta_j} e^{-a(\mathbf{x}_{j-1})\rho_j} a_{r_j}(\mathbf{x}_{j-1}) \, \mathrm{d}\rho_j \\
&= \pi_0(\mathbf{x}_0) \prod_{j=1}^{d} \frac{a_{r_j}(\mathbf{x}_{j-1})}{a(\mathbf{x}_{j-1})} (e^{-a(\mathbf{x}_{j-1})\gamma_j} - e^{-a(\mathbf{x}_{j-1})\eta j})
\end{aligned} \tag{12}$$

where $\mathbf{x}_j = \mathbf{x}_0 + \sum_{i=1}^{j} (\boldsymbol{\nu}^+_{r_i} - \boldsymbol{\nu}^-_{r_i})$. (Observe that if $\mathbf{X}(0) = \mathbf{x}_0$ then $\mathbf{x}_j$ is the state of the system after the j-th jump).
Let

$$\mathcal{T} = \{\Gamma : \Gamma \text{ is a cylinder set of the form } (11)\}$$

be the class of all cylinder sets of $\mathbb{Z}_+^n \times ([0,\infty) \times \{1,\dots,v\})^d$ and let $\sigma(\mathcal{T})$ be the sigma algebra on $\mathbb{Z}_+^n \times ([0,\infty) \times \{1,\dots,v\})^d$ generated by the class $\mathcal{T}$. Then it is clear from (12) that π extends to a probability measure on the measure space $\left\{\mathbb{Z}_+^n \times ([0,\infty) \times \{1,\dots,v\})^d, \sigma(\mathcal{T})\right\}$ with initial distribution π_0. Equation (12) tells us that the density of $\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^d\right)$ is given by

$$p\left(\left(\mathbf{x}_0, \{(\rho_k, r_k)\}_{k=1}^d\right) | \mathcal{C}\right) = \pi_0(\mathbf{x}_0) \prod_{j=1}^{d} e^{-a(\mathbf{x}_{j-1})\rho_j} a_{r_j}(\mathbf{x}_{j-1}). \tag{13}$$

Fix a time T. Let $\mathcal{S}([0,T], \mathbb{Z}_+^n)$ be the space of all piecewise-constant functions $f : [0,T] \to \mathbb{Z}_+^n$. Notice that a sample path of $\mathbf{X}$ up to time T lies in $\mathcal{S}([0,T], \mathbb{Z}_+^n)$. If $\mathbf{y} \in \mathcal{S}([0,T], \mathbb{Z}_+^n)$, then putting $t_0(\mathbf{y}) = 0$, define the jump times inductively by

$$t_k(\mathbf{y}) = \inf\left\{s > t_{k-1}(\mathbf{y}) | \mathbf{y}(s) \neq \mathbf{y}(t_{k-1}(\mathbf{y}))\right\}. \tag{14}$$

DOI: 10.1002/rnc

Moreover, let $r_k(\mathbf{y}) = \arg\{i|\mathbf{y}(t_k(\mathbf{y})) = \mathbf{y}(t_{k-1}(\mathbf{y})) + \boldsymbol{\nu}_i^+ - \boldsymbol{\nu}_i^-\}$ denote the reaction that occurred at $t_k(\mathbf{y})$. Suppose that $M = \sup\{k|t_k(\mathbf{y}) \leq T\}$ is the total number of jumps of $\mathbf{y}$ in the time interval $[0, T]$. Then the probability density of $\left(\mathbf{X}(0), \{(\tau_k - \tau_{k-1}, R_k)\}_{k=1}^M\right)$ evaluated at $\left(\mathbf{y}(0), \{(t_k - t_{k-1}, r_k)\}_{k=1}^M\right)$, $p(\mathbf{y}|\mathcal{C})$, is given by

$$p(\mathbf{y}|\mathcal{C}) = \pi_0(\mathbf{y}(0)) \prod_{j=1}^{M} \exp\left\{-a\left(\mathbf{y}(t_{j-1})\right)(t_j - t_{j-1})\right\} \times a_{r_j}\left\{\mathbf{y}(t_{j-1})\right\}. \tag{15}$$

Here for notational convenience, we have put $t_k \equiv t_k(\mathbf{y})$ and $r_k \equiv r_k(\mathbf{y})$.

Similarly, if $\mathbf{Z}(\cdot, \mathbf{b})$ satisfies (4) and $\mathbf{y}(\cdot, \mathbf{b})$ is a sample path realization of $\mathbf{Z}(\cdot, \mathbf{b})$ upto time T, then the corresponding $p(\mathbf{y}|\mathcal{C}, \mathbf{b})$ is given by

$$p(\mathbf{y}|\mathcal{C}, \mathbf{b}) = \pi_0(\mathbf{y}(0, \mathbf{b})) \prod_{j=1}^{M} \exp\left\{-\tilde{a}\left[\mathbf{y}(t_{j-1}, \mathbf{b}), \mathbf{b}\right](t_j - t_{j-1})\right\} \tilde{a}_{r_j}\left[\mathbf{y}(t_{j-1}, \mathbf{b}), \mathbf{b}\right]. \tag{16}$$

Notice that by the definition, it follows that the likelihood $L(\mathcal{C}, \mathbf{b}|\mathbf{y}) = p(\mathbf{y}|\mathcal{C}, \mathbf{b})$, where $p(\mathbf{y}|\mathcal{C}, \mathbf{b})$ is given by (16).

In our case the $\tilde{a}_k$ are given by (5). Consequently partitioning the observation such that $\mathbf{y} = (\tilde{\mathbf{y}}; \bar{\mathbf{y}})$ and using (8) we have

$$\begin{aligned} p(\tilde{\mathbf{y}}|\mathbf{b}, \mathcal{C}) = {} & \pi_0\left((\tilde{\mathbf{y}}(0, \mathbf{b}), \mathbf{F}(\tilde{\mathbf{y}}(0, \mathbf{b}), \mathbf{b}))\right) \\ & \times \prod_{j=1}^{M} \exp\left\{-a\left[(\tilde{\mathbf{y}}(t_{j-1}, \mathbf{b}), \mathbf{F}(\tilde{\mathbf{y}}(t_{j-1}, \mathbf{b}), \mathbf{b}))\right](t_j - t_{j-1})\right\} \\ & \times a_{r_j}\left[(\tilde{\mathbf{y}}(t_{j-1}, \mathbf{b}), \mathbf{F}(\tilde{\mathbf{y}}(t_{j-1}, \mathbf{b}), \mathbf{b}))\right], \end{aligned} \tag{17}$$

where $\bar{\mathbf{y}}$ denotes the components of $\mathbf{y}$ that are expressed as a function of $\bar{\mathbf{y}}$ using conservation laws. Both Bayesian as well as *frequentist* approaches involve evaluation of (15) to obtain estimates for $\mathcal{C}$. Even closed form solutions exist for this problem as it was shown in [25]. However, the heterogeneous case requires evaluation of equation (17), where we can observe a nonlinear dependency between $p(\tilde{\mathbf{y}}|\mathbf{b}, \mathcal{C})$ and $\mathbf{b}$ such that an analytical evaluation becomes intractable.

3.2. A Hierarchical Bayesian Model

Assuming that the we have measured a set of m independent paths $\mathcal{Y} = \left\{\tilde{\mathbf{y}}^1, \ldots, \tilde{\mathbf{y}}^m\right\}$ governed by the set of (unknown) conservation constants $\mathcal{B} = \left\{\mathbf{b}^1, \ldots, \mathbf{b}^m\right\}$, the conditional density over all paths is given by

$$p(\mathcal{Y}|\mathcal{C}, \mathcal{B}) = \prod_{i=1}^{m} p(\tilde{\mathbf{y}}^i|\mathcal{C}, \mathbf{b}^i). \tag{18}$$

Next, prior distributions over the unknown quantities are introduced. For instance, any a priori knowledge - maybe obtained from previous experiments - should be incorporated at this point to reduce the estimator's variance. In the following, priors over the chemical rate constants are denoted as $p(\mathcal{C})$.

Furthermore, we introduce priors for the conservation constants $p(\mathbf{b}^i), \forall i \in \{1, \ldots, m\}$. In particular, we model each of the parameters $\mathbf{b}^i$ as random vectors, which are statistically independent from each other but drawn from one common distribution $p(\mathbf{b}^i|\boldsymbol{\alpha}) = p(\mathbf{b}|\boldsymbol{\alpha}), i \in \{1, \ldots, m\}$, where $\boldsymbol{\alpha}$ denotes a set of hyperparameters, governing position and shape of that distribution. Clearly, additionally estimating $\boldsymbol{\alpha}$ from the data will give useful information regarding the statistics of the cell variability. We can write the joint density function over all variables as

DOI: 10.1002/rnc

$$p(\mathcal{Y}, \mathcal{C}, \mathcal{B}, \boldsymbol{\alpha}) = \left\{ \prod_{i=1}^{m} p(\tilde{\mathbf{y}}^i | \mathbf{b}^i, \mathcal{C}) p(\mathbf{b}^i | \boldsymbol{\alpha}) \right\} p(\mathcal{C}) p(\boldsymbol{\alpha}) \tag{19}$$

and the corresponding graphical model is depicted in Figure 2.

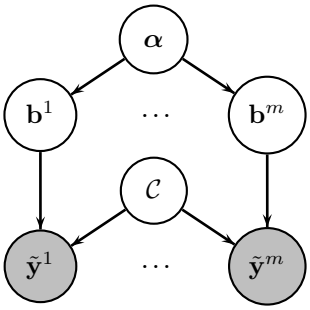


Figure 2. Graphical model for estimating the reaction rate constants $\mathcal{C}$, the conservation constants $\mathbf{b}^1, \ldots, \mathbf{b}^m$ as well as their hyperparameters $\boldsymbol{\alpha}$. Shaded circles denote observed quantities.

Given (19), we can calculate the posterior distribution $p(\mathcal{C}, \mathcal{B}, \boldsymbol{\alpha} | \mathcal{Y})$ using Bayes' theorem. Note that for each observed path, another cell-specific conservation constant $\mathbf{b}^i, \forall i \in \{1, \ldots, m\}$ enters the model, which introduces additional uncertainty accordingly. For this reason, we prefer calculating the marginal density

$$p(\mathcal{C}, \boldsymbol{\alpha} | \mathcal{Y}) = \int p(\mathcal{C}, \mathcal{B}, \boldsymbol{\alpha} | \mathcal{Y}) d\mathcal{B}, \tag{20}$$

where the hidden variable set $\mathcal{B}$ is integrated out and thus, only $\boldsymbol{\alpha}$ has to be estimated in order to robustly capture variability due to extrinsic noise. As already mentioned, it turns out that (20) is analytically complex and thus, suitable approximation schemes have to be used. In this work, we apply *Markov chain Monte Carlo* (*MCMC*) [26] methods, to draw samples from the target posterior.

4. A CASE STUDY: HOG1 MAPK PROTEIN EXPRESSION

In this section, the proposed algorithms from Section 3 are applied to a reaction system, modeling the Hog1 driven transcriptional process in yeast cells. The MAPK (Mitogen Activated Protein Kinase) Hog1 is the most downstream kinase of a signaling cascade which is activated by osmotic stress [27]. Upon activation of this pathway, a large fraction of the activated Hog1 is relocated from the cytoplasm to the nucleus of the cell to initiate a transcriptional program resulting in the up-regulation of roughly 300 genes [28]. The kinase activity of Hog1 in the cytoplasm leads to the production of glycerol, which allows the cells to equilibrate the interior and exterior osmotic pressures of the cell. Once adaptation has been achieved, Hog1 activity returns to basal levels and the active MAPK leaves the nucleus, as shown in Figure 3(a), thereby offering only a short temporal window for the transcription of downstream target genes. This transient relocation of Hog1 measured by microscopy will be used as an input for our simulation [29].

4.1. The Model

We consider a variant of the model proposed in [6], which focuses on the activation of a single gene upon entry of the active MAPK $HOG1^P$ in the nucleus. This species can bind via a transcription factor to the gene coding for the $STL1$ protein $GENE^{STL1}$. This complex $HOG1^P : GENE^{STL1}$

DOI: 10.1002/rnc

will generate newly synthesized proteins $STL1$. This protein production rate can however be enhanced if chromatin remodeling factors such as the RSC complex can interact with the transcription initiation complex to open the chromatin structure and allow for faster transcription of the gene product [30]. We therefore assume that c_8 is much larger than c_3. $HOG1^P$ can be dephosphorylated via the nuclear phosphatase $PTP2$ and generate an inactive form of the enzyme $HOG1$.

The described model can be written as

$$\begin{aligned}
HOG1^P + GENE^{STL1} &\underset{c_2}{\overset{c_1}{\rightleftharpoons}} HOG1^P : GENE^{STL1} \\
HOG1^P : GENE^{STL1} &\xrightarrow{c_3} HOG1^P : GENE^{STL1} + STL1 \\
HOG1^P + RSC &\underset{c_5}{\overset{c_4}{\rightleftharpoons}} HOG1^P : RSC \\
HOG1^P : RSC + GENE^{STL1} &\underset{c_7}{\overset{c_6}{\rightleftharpoons}} HOG1^P : RSC : GENE^{STL1} \\
HOG1^P : RSC : GENE^{STL1} &\xrightarrow{c_8} HOG1^P : RSC : GENE^{STL1} + STL1 \\
STL1 &\xrightarrow{c_9} 0 \\
HOG1^P + PTP2 &\underset{c_{11}}{\overset{c_{10}}{\rightleftharpoons}} HOG1^P : PTP2 \\
HOG1^P : PTP2 &\xrightarrow{c_{12}} HOG1 + PTP2,
\end{aligned} \tag{21}$$

where all reactions are assumed to follow mass-action kinetics. In reference to the introductory Section 1 we remark that the synthesis reactions of protein $STL1$ are clearly not elementary mechanistic events. For this high level of aggregation, extrinsic noise can enter the synthesis rates c_3 and c_8. For instance the number of ribosomes that is as well subsumed in this aggregated synthesis reaction varies from cell-to-cell. Although, such distributions over kinetic parameters can easily be incorporated into the proposed framework, we refrain from doing so here for the sake of clarity. With the system state[§]

$$\mathbf{X}(t) = \begin{pmatrix} HOG1^P \\ GENE^{STL1} \\ RSC \\ STL1 \\ HOG1^P : GENE^{STL1} \\ HOG1^P : RSC \\ HOG1^P : RSC : GENE^{STL1} \\ PTP2 \\ HOG1^P : PTP2 \\ HOG1 \end{pmatrix} \tag{22}$$

and the reaction system from (21), we can identify four conservation laws, i.e., a four-dimensional null space

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \mathbf{X}(t) = \begin{pmatrix} \tilde{\mathbf{b}} \\ \mathbf{b} \end{pmatrix}.$$

[§]Additionally, the time-dependence of the species is not explicitly written in the following.

DOI: 10.1002/rnc

Note that $HOG1^P$ is treated as an input signal to our model and thus, conservation will be violated as soon as $HOG1^P$ molecules enter or leave the nucleus. Furthermore, there exists only one $GENE^{STL1}$ sequence in the genome, such that cell-to-cell variability will not affect this quantity. Thus, only the two conservation laws corresponding to $\mathbf{b} = (b_1, b_2)^T$ are used to model extrinsic variability, i.e.,

$$RSC + HOG1^P : RSC + HOG1^P : RSC : GENE^{STL1} = b_1 \tag{23}$$

$$PTP2 + HOG1^P : PTP2 = b_2. \tag{24}$$

In the following, $\mathbf{b}^i = (b_1^i, b_2^i)^T$ denotes the conservation constant corresponding to the i-th cell.

With the previous definitions, all variations due to extrinsic noise are expressed in the conservation constants $\mathbf{b}^i$ in a condensed manner. To obtain a functional dependency between the conditional path density and $\mathbf{b}^i$, we make use of equation (8)¶. As a consequence, we can express two species by the other ones, knowing $\mathbf{b}^i$. For the following experiments we chose $HOG1^P : RSC$ as well as $HOG1^P : PTP2$ to be expressed by relation (8). As the two conservation laws from our model do not involve common species, we obtain the following simple expression:

$$\begin{pmatrix} HOG1^P : RSC \\ HOG1^P : PTP2 \end{pmatrix}^i = \mathbf{b}^i - \begin{pmatrix} RSC + HOG1^P : RSC : GENE^{STL1} \\ PTP2 \end{pmatrix}^i, \tag{25}$$

where superscript i again denotes the i-th cell. As (25) allows us to rewrite the conditional path density in terms of $\mathbf{b}^i$, we can jointly estimate the quantities of interest (i.e., $\mathcal{C}$ and $\boldsymbol{\alpha}$) according to the Bayesian model of Section 3.2.

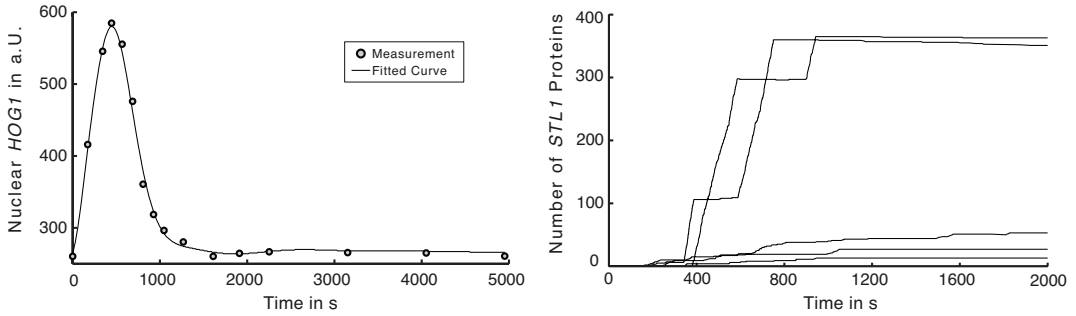
4.2. Experimental Setup

In this work, we evaluate the proposed algorithms using synthetic time-course data, generated by running Gillespie's direct simulation algorithm [31] for the model specification described below.
As already pointed out, the model from the previous section does not capture the activation of $HOG1^P$ due to osmotic stress as well as its deactivation, induced by a negative feedback mechanism relying on the intracellular accumulation of glycerol. Instead, we consider total nuclear Hog1 (i.e., $HOG1 + HOG1^P$) as an input signal obtained from real-world experimental data. Hog1 relocation was measured by tagging this protein with a yellow fluorescent protein (YFP). The localization of Hog1-YFP was followed by time-lapse microscopy in conjunction with the nuclear signal obtained from a CFP-tagged histone. Segmentation of the CFP and YFP images allows to measure the nuclear accumulation of Hog1 as a function of time [29]. To increase the resolution of the sparse measurement data, a linear regression was performed using radial basis functions [32]. Figure 3(a) shows the measured data obtained after adding $0.1M$ of $NaCl$ to the media as well as the fitted nuclear Hog1 response.

Clearly, many other genes are target of $HOG1^P$ and the actual number of $HOG1^P$ proteins available for the $STL1$ promoter will be a fraction thereof. That is accounted for by rescaling and rounding the analytical function to get copy numbers on the interval $[0, 60]$. To incorporate the external data into Gillespie's simulation algorithm, the change of nuclear Hog1 molecules Δ_{Hog1} - governed by the fitted time function - is cumulated between two subsequent reactions r_j and r_{j+1} and then added to the current state of $HOG1^P$ if $\Delta_{Hog1} > 0$ or the current state of $HOG1$ if $\Delta_{Hog1} < 0$ (which means that $HOG1$ decreases). Note that this is based on the assumption that Hog1 can enter the nucleus only in it's active form $HOG1^P$ and only inactive Hog1 (i.e., $HOG1$) is exported. This procedure makes sure that the resulting inhomogeneity does not affect the stochastic simulation, as the Hog1 - changes happen exactly at the jump times t_k and thus, the propensities remain constant between two subsequent reactions [33].
The stochastic rates of the reference model where chosen such as to reproduce a bimodality in the

¶Note that (8) is still valid if only a subset of the conservation laws is used.

DOI: 10.1002/rnc



(a) (b)

Figure 3. Total nuclear Hog1 measurement and $STL1$ protein expression. 3(a): Estimated Hog1 response after adding $0.1M$ of $NaCl$. The resulting function was obtained using a ridge regression with radial basis functions (RBFs). The RBF kernels were placed directly on the acquired measurement points and the width-parameter of the kernels was adjusted according to the increasing measurement intervals. 3(b): Time-course data for $STL1$, obtained from the described reference model for 5 different cells. Some of the cells start producing $STL1$ much faster than the other ones. This gives rise to the bimodality in $STL1$ protein expression.

expression of $STL1$ [6]. The interplay between the transient Hog1 activation pattern and a slow remodeling of the transcribed gene by chromatin can result in a split in the cell population. In a fraction of the runs, the RSC complex will be able to act on the gene expression mechanism and it will result in high expression of the target protein. In the other fraction of the simulations, protein production will rely solely on the slow basal protein expression rate, which will result in a low level of the protein. Table I shows the stochastic rates for the reference system, whereas quantities marked with * are estimated during the model calibration (i.e., $\mathcal{C}^* = \{c_3, c_6, c_8, c_{11}, c_{12}\}$).

Table I. Stochastic rates of the reference model.

Name	Value in s^{-1}	Name	Value in s^{-1}
c_1	0.0010	c_7	0.01000
c_2	0.0400	c_8^*	1.50000
c_3^*	0.1000	c_9	0.00001
c_4	0.0005	c_{10}	0.00050
c_5	0.0600	c_{11}^*	0.01000
c_6^*	0.0003	c_{12}^*	0.03000

To simulate the cell-to-cell variability, we assume that for each cell $i \in \{1, \ldots, m\}$, the conservation constants $\mathbf{b}^i$ are log-normally [34] distributed[‖] according to

$$\begin{pmatrix} b_1^i \\ b_2^i \end{pmatrix} \sim \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{26}$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \ln 80 \\ \ln 20 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} 1.2 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}.$$

[‖] As our stochastic model deals with discrete copy numbers, the values for RSC as well $PTP2$ have to be rounded after drawing them from a continuous distribution. This procedure introduces nonlinear artifacts, which are assumed to be negligibly small.

DOI: 10.1002/rnc

The initial multiplicities of the complexes $HOG1^P : RSC_{t=0}$, $HOG1^P : RSC : GENE^{STL1}_{t=0}$ as well as $HOG1^P : PTP2_{t=0}$ are chosen to be zero such that the extrinsic noise is entirely reflected by the initial amount of RSC and $PTP2$. Note that we assume constant cell volume for this case study. The variability in abundance is thus due to a variability in concentration. A complete list of the species and their initial copy numbers can be found in Table II.

Table II. Species and initial copy numbers.

Name	Initial Copy Number
$HOG1^P$	0
$GENE^{STL1}$	1
RSC	See eq. 26
$STL1$	0
$HOG1^P : GENE^{STL1}$	0
$HOG1^P : RSC$	0
$HOG1^P : RSC : GENE^{STL1}$	0
$PTP2$	See eq. 26
$HOG1^P : PTP2$	0
$HOG1$	20

Figure 3(b) shows several sample paths for $STL1$ protein expression using the described model configuration.

4.3. Model Calibration

A generic description of our Bayesian model used for the following inference task was already given in Section 3. For a better illustration, we have incorporated our assumptions regarding the experimental setup from the previous Sections 4.1 and 4.2 into a problem-specific graphical model, depicted in Figure 4. One important aspect that can be nicely seen from Figure 4 is that the hyperparameter $\boldsymbol{\Sigma}$ also allows us to capture correlations between different components of the conservation constant $\mathbf{b}^i$, i.e., b_1^i and b_2^i. In other words, it is often plausible to assume that those quantities will not change independently due to extrinsic noise sources, which would be reflected by the off-diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$.

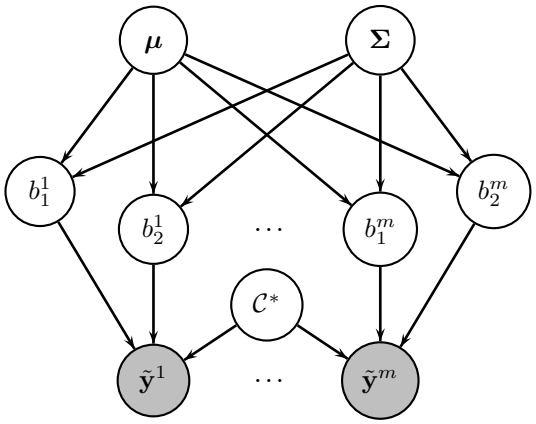


Figure 4. Graphical model for calibrating our $HOG1^P$-model. Observe that each conservation constant depends on the covariance matrix $\boldsymbol{\Sigma}$, allowing us to capture correlations between them.

In order to calibrate our specific model to the reference data, we need to evaluate the marginal posterior distribution

DOI: 10.1002/rnc

$$p(\mathcal{C}^*, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathcal{Y}) = \int p(\mathcal{C}^*, \mathcal{B}, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathcal{Y}) d\mathcal{B}. \tag{27}$$

Unfortunately, the quantities of interest enter the posterior in a non-linear manner, which makes exact inference impossible. To circumvent this problem, one could for instance approach approximate inference principles such as variational Bayesian techniques [35, 16]. A simple but powerful alternative are sampling-based methods, such as MCMC [26], which allow drawing samples from the distribution of interest, in order to obtain a numerical representation - for instance - of our posterior density from (27). As we will see shortly, also marginalization with respect to some of the parameters is trivial using MCMC samplers. In this work, we will evaluate the posterior distribution using a Metropolis-Hastings-type sampler [26], as it is easy-to-use and it only requires the joint density function to be evaluable (which is the case for our model). Unfortunately, it turns out that especially for complex and high-dimensional posterior distributions, the standard Metropolis-Hastings algorithm achieves only poor acceptance rates, which leads to a huge computational effort for retrieving good density estimates. The need for a computationally efficient sampling scheme suggests using *block-at-a-time* Metropolis-Hastings (*BATMH*) samplers [36, 26]. Using BATMH algorithms, the parameter set $\mathcal{P}$ of the the target posterior $p(\mathcal{P}|\mathcal{Y})$ is divided into disjoint blocks $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_u$, i.e., $\mathcal{P} = \cup_{k=0}^{u} \mathcal{P}_k$ and $\mathcal{P}_k \cap \mathcal{P}_l = \emptyset, \forall k \neq l$. Then - like in standard Metropolis-Hastings - new samples $\hat{\mathcal{P}}_k^{new}$ are proposed but the decision on the acceptance is performed for each of the blocks separately, using the most recent accepted values of all the other blocks $\hat{\mathcal{P}}_{l\neq k}^{old}$. Mathematically, this means that a block proposal $\hat{\mathcal{P}}_k^{new}$ is accepted with probability

$$a_k = \min\left\{1, \frac{q_k(\hat{\mathcal{P}}_k^{old}|\hat{\mathcal{P}}_k^{new}) p(\mathcal{Y}, \hat{\mathcal{P}}_k^{new}|\hat{\mathcal{P}}_{l\neq k}^{old})}{q_k(\hat{\mathcal{P}}_k^{new}|\hat{\mathcal{P}}_k^{old}) p(\mathcal{Y}, \hat{\mathcal{P}}_k^{old}|\hat{\mathcal{P}}_{l\neq k}^{old})}\right\}, \tag{28}$$

where $q_k(.|.)$ denotes the proposal density for block k. To keep the notation compact, we use $q(.|.)$ as a proxy for the different proposals $q_k(.|.)$ in the following, but the reader should keep in mind that they are not necessarily equal for each block $\mathcal{P}_k$. Using BATMH-methods, acceptance rates can often be drastically improved, as it is typically easier to design good proposal densities for each of the blocks separately instead of for the entire parameter set [37]. In the following few lines, we will give several details on the configuration of the applied BATMH scheme.

First of all, the overall parameter set $\mathcal{P} = \{\mathcal{C}^*, \mathcal{B}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ is divided into $m+2$ blocks such that $\mathcal{P}_i = \left\{\mathbf{b}^i\right\}, i \in \{1, \ldots, m\}$, furthermore $\mathcal{P}_{m+1} = \{\mathcal{C}^*\}$ and $\mathcal{P}_{m+2} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. Note that for each of the m cells, the dimensionality of the posterior increases by the number of conservation laws used, which would in general lead to very low acceptance probabilities as soon as many cells are considered. By introducing one individual block for each of the cell-specific conservation constants $\mathbf{b}^i$, we can break up the dependency between the acceptance probability and the number of cells used for calibration. Additionally, we can easily find good proposals for each of the blocks, conditioned on the parameters from the other blocks. For the conservation constants b_1^i and b_2^i we use discrete uniform distributions of width $L = 10$, centered around the previous accepted value, i.e., $q(b^{new}|b^{old}) = \mathcal{U}\left(b^{old} - 5, b^{old} + 5\right)$. As the rate constants usually range over several orders of magnitude, it makes sense to reflect this by the proposal distribution. In our particular case, log-normal distributions have achieved very good acceptance rates, whereas the logarithmic mean was set to the logarithm of the previous accepted rate value and the variance in the log-domain was set to 0.03^2, i.e., $q(c^{new}|c^{old}) = \mathcal{LN}\left(\ln c^{old}, 0.03^2\right)$. Furthermore, the components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ were sampled using one-dimensional normal distributions, i.e., $q(\mu^{new}|\mu^{old}) = \mathcal{N}\left(\mu^{old}, 0.3^2\right)$ for μ_1 and μ_2 and $q(\Sigma^{new}|\Sigma^{old}) = \mathcal{N}\left(\Sigma^{old}, 0.02^2\right)$ for Σ_{11}, Σ_{22} and Σ_{12} respectively.

Note that when using the described sampling algorithm, priors for the rate constants $\mathcal{C}^*$ could be easily incorporated. However, we decided not to incorporate prior knowledge, allowing us to objectively compare the resulting estimates.

After running the described BATMH-scheme until convergence, the resulting sequence of parameter values could be used to estimate the target density $p(\mathcal{P}|\mathcal{Y})$. However, in many cases it is not

DOI: 10.1002/rnc

necessary to sample from the full posterior. Instead, it is often sufficient to obtain density estimates of the marginal distributions for some parameter subsets. In fact, working with the marginal posteriors can even simplify the interpretation and evaluation of the obtained results. Using MCMC methods, marginal posteriors can be calculated by "projecting out" certain dimensions of the Markov chain, which means that those dimensions are simply omitted during the histogram computation. In this work, we are interested in calculating the marginals $p(c_i|\mathcal{Y}), i \in \{3, 6, 8, 11, 12\}, p(\mu_1|\mathcal{Y}), p(\mu_2|\mathcal{Y})$ as well as $p(\Sigma_{11}|\mathcal{Y}), p(\Sigma_{22}|\mathcal{Y})$ and $p(\Sigma_{12}|\mathcal{Y})$.

4.4. Results

We have performed the model estimation task by applying the BATMH algorithm from above using $m = 50$ cell observations generated with Gillespie's stochastic simulation algorithm. Such a number is in the right orders of magnitude for today's fluorescent microscopy setup. In fact, the experimental platform can provide recordings of 100 cells per time point by taking sequential images at different location of the microfluidic imaging chamber.
Each of the traces was simulated up to a maximum time horizon of $T = 2000s$. The BATMH sampler was used to create a Markov chain of length $M = 150000$, whereas each parameter block was updated once per iteration. The first 60000 parameter values have been discarded to avoid biasing the results by the algorithm's "burn in". The initial values used for the BATMH sampler are summarized in Table III.

Table III. Parameter values used for initialization of the BATMH sampler.

Name	Parameter Value
$c \in \mathcal{C}^*$	$0.01s^{-1}$
b_1^i	1000
b_2^i	500
μ_1	$\ln 1000$
μ_2	$\ln 500$
Σ_{11}	1
Σ_{22}	1
Σ_{12}	0

The running time of the sampling scheme under the described configuration was approximately 20 hours using a MATLAB (2010b, The MathWorks, Natick, MA) implementation on a standard personal computer. Figures 5 and 6 show the estimation results for the marginal posteriors for the rate constants $\mathcal{C}^*$ as well as the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively. Each of the histograms was calculated using 20 bins which have additionally been rescaled such as to approximate the underlying continuous density**. The solid lines indicate the true value of the respective quantity, illustrating the estimators accuracy. For the hyperparameters in Figure 6 we have also computed maximum likelihood estimates (MLE, dashed lines) based on the m conservation constants $\mathbf{b}^i$ that were drawn from the reference distribution (26). Note that these values represent the best possible estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ knowing $\mathbf{b}^i, i \in \{1, \ldots, m\}$ and if no prior knowledge is available [38]. Furthermore, we have computed the empirical mean values over the 90000 parameter samples, representing minimum mean square error (*MMSE*, [38]) estimates of the model quantities. The results are summarized in Table IV.

Thus, we are now ready to perform a comparison between the reference system and the calibrated system and in order to do so, we computed histograms for the copy number of $STL1$ at time $T = 2000s$ using $m = 50000$ sample paths for both systems. As we have also computed estimates of the hyperparameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, acting as a proxy for the extrinsic noise, we can eliminate cell-to-cell

**Thus, they integrate up to one.

DOI: 10.1002/rnc

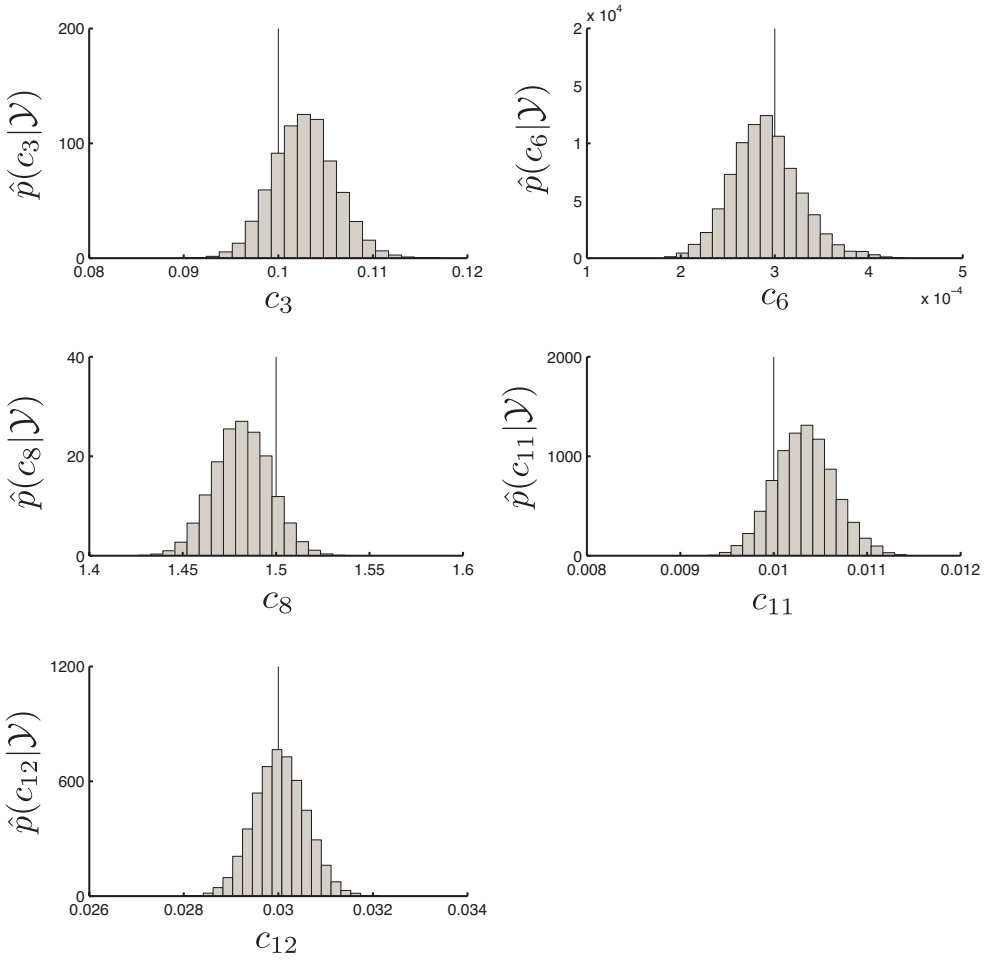


Figure 5. Estimated marginal posteriors for the stochastic rate constants c_3, c_6, c_8, c_{11}, c_{12}. The reference values are indicated by vertical lines.

Table IV. Minimum mean square error (MMSE) estimates of the unknown model quantities.

Quantity	True Value	Estimate
c_3	$0.1000s^{-1}$	$0.1027s^{-1}$
c_6	$0.0003s^{-1}$	$0.0003s^{-1}$
c_8	$1.5000s^{-1}$	$1.4817s^{-1}$
c_{11}	$0.0100s^{-1}$	$0.0103s^{-1}$
c_{12}	$0.0300s^{-1}$	$0.0300s^{-1}$
μ_1	4.3820	4.2351
μ_2	2.9957	2.9813
Σ_{11}	1.2	1.3927
Σ_{22}	0.4	0.4676
Σ_{12}	0.6	0.7256

variations by setting all conservations constants $\mathbf{b}^i$ to the expected value of the density $p(\mathbf{b}^i|\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \mathcal{LN}\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$, with $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ as the MMSE estimates from Table IV. Figure 7 depicts logarithmic histograms of $STL1$ for the reference system, the calibrated system as well as the calibrated

Prepared using rncauth.cls
DOI: 10.1002/rnc

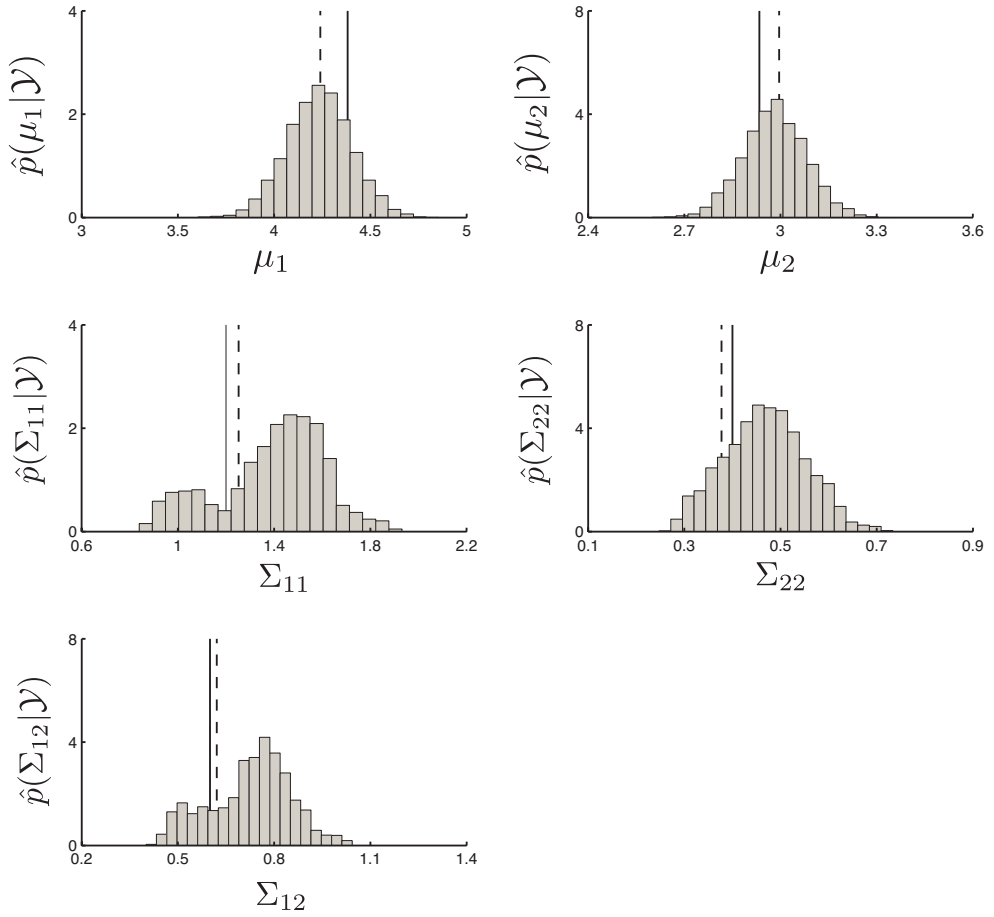


Figure 6. Estimated marginal posteriors for the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The solid vertical lines indicate the reference values. For comparison, also the maximum likelihood estimate (MLE) was calculated using the true conservation constants $\mathbf{b}^i$ (dashed line).

system with the extrinsic variability "switched off", eliminating the heterogeneity introduced by the extrinsic noise. In general, the variability in mass-conservation $\mathbf{b}^i$ can lead to a significant "smearing" of the expression profiles. Note that this effect is limited in our case, as the species RSC and $PTP2$ are counteracting regarding the protein expression. Increasing RSC will end up with a faster production of $STL1$, while $PTP2$ dephosphorylates $HOG1^P$ and thus controls the length of the temporal window in which $HOG1^P$ can bind to the promoter. As we assume the corresponding conservation constants b_1^i and b_2^i to be positively correlated in (26), the smearing is limited. Nevertheless, the extrinsic variability has a significant impact on the shape of the $STL1$ distribution as can be seen in Figure 7. Note that the performed *in silico* cell homogenization gives rise to an average cell, with all the variability coming from intrinsic stochasticity. Interestingly, subject to the stochastic dynamics such typical cells apparently have a higher change for efficient gene expression by RSC recruitment than the heterogeneous population of cells. In particular, their mean expression is higher than that of their heterogeneous counterpart – indicating clearly that the averaging does not commute with the dynamical system.

DOI: 10.1002/rnc

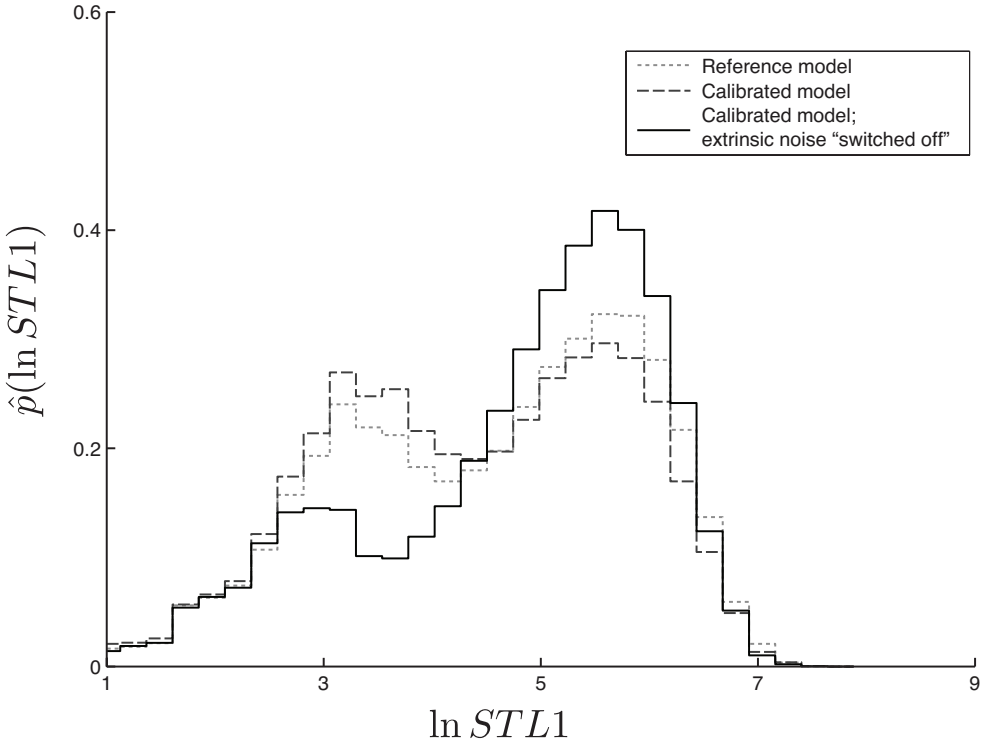


Figure 7. $STL1$ profile recorded after $T = 2000$ seconds for $m = 50000$ cells. The results demonstrate that the expression profile is slightly smoothed out due to extrinsic noise (gray dashed and green line). The red line corresponds to the calibrated system whereby the extrinsic noise was "switched off", i.e. the population was homogenized.

5. CONCLUSION

We propose a novel unified framework to address cell-to-cell variability observed in single-cell data. It accounts for extrinsic noise or heterogeneity and intrinsic noise caused by the stochastic nature of chemical kinetics. The extrinsic variability is captured by allowing cells to have different total copy-numbers of particular species. This novel approach is in contrast to previous attempts to reflect cell-to-cell variability solely through variations in kinetic rate constants. We argue that kinetic rate constants of mechanistic events should not be subject to variations within a population of isogenic cells. Copy-number variations can cause variations in rate parameters of aggregated rate laws and thus the proposed framework provides a more principled approach to generate such parametric variations. We remark however, that in our context of stochastic chemical kinetics, the usage of aggregated rate laws is questionable *per se* as all common aggregations are based on time-scale considerations performed on the reaction-rate equation and thus lack stochastic validity.

We formulate a heterogeneous version of the underlying Markov jump process using the random time change model. For clarity of the exposition we chose to introduce the generic framework by assuming fully observed sample paths and no unobserved species in the reaction subspace. Based on that assumption, the conditional path densities that are required for a proper Bayesian treatment are rigorously obtained. The estimation scheme rests upon a hierarchical Bayesian model. Thus, we consider hyperparameters controlling the variability of the mass-conservation constraints. As the path densities depend nonlinearly on the varying total protein concentration, the analytical result of [22] does not apply. To sample the generally high-dimensional posterior we propose Markov-chain-Monte-Carlo methods.

The framework is illustrated using a small model for the osmo-stress induced transcription in budding yeast. It is a 10-dimensional system with 12 reactions involving only mass-action kinetics.

DOI: 10.1002/rnc

We assume 5 out of the 12 kinetic parameters to be unkown and provide minimum mean square error estimates. Jointly with them, 5 hyperparameters are estimated, that control the variability and correlation of and among mass-conserved species. The sampling is done using an advanced block-at-a-time Metropolis-Hasting scheme in order to obtain good acceptance ratios. After showing the validity of this approach here, our future work will involve the incorporation of this method into a general estimation scheme, allowing for discretely observed time-series data and unobserved species.

ACKNOWLEDGEMENTS

S.P. and M.P. acknowledge the support from the EU UNICELLSYS project and the Swiss SystemsX.ch organization and the ETHZ/UNIZ Competence Centre Systems Physiology and Metabolic Disease (CC-SPMD).

REFERENCES

1. Altschuler SJ, Wu LF. Cellular heterogeneity: Do differences make a difference. *Cell* 2010; **141**:559–563.
2. Rao C, Wolf D, Arkin A. Control, exploitation and tolerance of intracellular noise. *Nature* 2002; **420**:231–237.
3. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science* 2002; **297**:1183–1186.
4. Blake WJ, Kearns N, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature* 2003; **442**:633–637.
5. Paliwal S, Iglesias P, Campell K, Hilioti Z, Groisman A, Levchenko A. MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature* 2007; **446**:46–51.
6. Pelet S, Rudolf F, Nadal-Ribelles M, de Nadal E, Posas F, Peter M. Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 2011; **332**(6030):732–735, doi:10.1126/science.1198851.
7. McQuarrie DA. Stochastic approach to chemical kinetics. *J. Appl. Probab.* 1967; **4**(3):413–478.
8. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce G, Brent R. Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 2005; **437**(29):699–502.
9. Raser JM, O'Shear EK. Control of stochasticity in eukaryotic gene expression. *Science* 2004; **304**:1811–1814.
10. Yanagida T, Ishii Y ((eds.)). *Single-molecule dynamics in life science*. Wiley-VCH: Weinheim, 2009.
11. Raj A, van Oudenaarden A. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* 2009; **38**:255–270.
12. Boys RJ, Wilkinson DJ, Kirkwood TBL. Bayesian inference for a discretely observed stochastic kinetic model. *Stat. Comput.* 2008; **18**(2):125–135.
13. Reinker S, Altman RM, Timmer J. Parameter estimation in stochastic biochemical reactions. *IEE Proc.-Syst. Biol.* 2006; **153**(4):168–178.
14. Golightly A, Wilkinson DJ. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Stat. Data An.* 2008; **83**:1891–1901.
15. Wang Y, Christley S, Mjolsness E, Xie X. Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent. *BMC Systems Biology* 2010; **4**(1):99, doi:10.1186/1752-0509-4-99.
16. Ruttor A, Sanguinetti G, Opper M. Approximate inference for stochastic reaction processes. *Learning and Inference in Computational Systems Biology*. The MIT Press, 2009; 189–205.
17. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. Roy. Soc. Interface* 2009; **6**(31):187–2022.
18. Rachev ST. *Probability metrics and the stability of stochastic models*. Probability and Mathematical Statistics, Wiley: New York, 1991.
19. Thorsley D, Klavins E. Model reduction of stochastic processes using Wasserstein pseudometrics. *Amer Contr Conf*, 2008; 1374–1381.
20. Koeppl H, Setti G, Mangia M, Petrov T, Pelet S, Peter M. Probability metrics to calibrate stochastic chemical kinetics. *IEEE Symposium on Circuits and Systems*, 2010; 541–544.
21. Hasenauer J, Waldherr S, Radde N, Doszczak M, Allgoewer F. A maximum likelihood estimator for parameter distributions in heterogeneous cell populations. *Procedia Computer Science*, vol. 1, 2010; 1655–1663.
22. Wilkinson DJ. *Stochastic Modelling for Systems Biology*. Chapman & Hall, 2006.
23. Anderson DF, Ganguly A, Kurtz TG. Error analysis of tau-leap simulation methods. *accepted to Annals of Applied Probability* 2011; URL `http://arxiv.org/abs/0909.4790`.
24. Anderson DF, Kurtz TG. Continuous time Markov chain models for chemical reaction networks. *in press* 2011; URL `http://www.math.wisc.edu/~kurtz/papers/AndKurJuly10.pdf`.
25. Wilkinson DJ. *Stochastic Modelling for Systems Biology (Chapman & Hall/CRC Mathematical & Computational Biology)*. 1 edn., Chapman and Hall/CRC, 2006.
26. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* April 1970; **57**(1):97–109, doi:10.1093/biomet/57.1.97.
27. Hohmann S, Krantz M, Nordlander B. Yeast osmoregulation. *Meth Enzymol* Jan 2007; **428**:29–45.
28. Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, O'Shea EK. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat Genet* Nov 2008; **40**(11):1300–6.

29. Schaber J, Adrover MA, Eriksson E, Pelet S, Petelenz-Kurdziel E, Klein D, Posas F, Goksör M, Peter M, Hohmann S, *et al*.. Biophysical properties of saccharomyces cerevisiae and their relationship with HOG pathway activation. *European biophysics journal : EBJ* Oct 2010; **39**(11):1547–56.
30. de Nadal E, Posas F. Multilayered control of gene expression by stress-activated protein kinases. *EMBO J* Jan 2010; **29**(1):4–13.
31. Gillespie DT. A rigorous derivation of the chemical master equation. *Physica A* 1992; **188**:404–425.
32. Bishop CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 1st ed. 2006. corr. 2nd printing edn., Springer, 2007.
33. Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *Journal of Chemical Physics* Dez 2007; **127**(21).
34. Furusawa C, Suzuki T, Kashiwagi A, Yomo T, Kaneko K. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophys*. 2005; **1**:25–31.
35. Beal MJ. Variational algorithms for approximate bayesian inference. PhD Thesis, Gatsby Computational Neuroscience Unit, University College London 2003.
36. Neath RC, Jones GL. Variable-at-a-time implementations of Metropolis-Hastings. *ArXiv e-prints* Mar 2009; .
37. Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *The American Statistician* Nov 1995; **49**(4):327–335.
38. Kay SM. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory (v. 1)*. 1 edn., Prentice Hall, 1993.

DOI: 10.1002/rnc